

**Scuola di Scienze Politiche**  
**Compito di Metodi Statistici per la Ricerca Sociale**  
**19 Dicembre 2018**

**Esercizio 1**

I seguenti dati si riferiscono al salario annuo lordo (in migliaia di Euro) di un campione di soggetti classificati per titolo di studio:

Gruppo	Numerosità del gruppo	Media Campionaria	Varianza Campionaria
$j$	$n_j$	$\bar{y}_j$	
Diploma	25	42.5	9
Laurea	20	45.0	10.24

Si supponga che il salario annuo lordo abbia una distribuzione Normale di media  $\mu_D$  e varianza  $\sigma_D^2$  tra i diplomati e Normale di media  $\mu_L$  e varianza  $\sigma_L^2$  tra i laureati. Si supponga inoltre che  $\sigma_D^2 = \sigma_L^2 = \sigma^2$

- Stimare la varianza  $\sigma^2$  comune alle due popolazioni
- Stimare la differenza tra le medie e il suo errore standard
- Verificare l'ipotesi che il salario medio sia lo stesso versus che sia diverso per laureati e diplomati. (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 10%; (iii) Calcolare il valore della statistica test e prendere una decisione
- Costruire un intervallo di confidenza al livello del 90%. Interpretare l'intervallo di confidenza ottenuto e metterlo in relazione con il risultato ottenuto al punto (c)

**Esercizio 2**

I seguenti dati sono relativi a un campione di 5 individui adulti su cui si sono rilevate le seguenti variabili:  $X$  = anni di istruzione e  $Y$  = qualità della vita (misurata su scala 0-100).

$$\sum_{i=1}^5 x_i = 80 \quad \sum_{i=1}^5 y_i = 350 \quad \sum_{i=1}^5 (x_i - \bar{x})^2 = 114 \quad \sum_{i=1}^5 (y_i - \bar{y})^2 = 840 \quad \sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = 285$$

- Stimare i coefficienti del modello di regressione che pone la qualità della vita in funzione del numero di anni di istruzione
- Stimare la varianza degli errori
- Verificare l'ipotesi che il coefficiente di regressione relativo alla variabile esplicativa anni di istruzione sia significativamente diverso da zero. (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Calcolare il valore della statistica test; (iii) Il  $p$ -valore per tale test è 0.026. Interpretare tale valore e prendere una decisione usando il  $p$ -value
- Calcolare la qualità della vita media per soggetti con 20 anni di istruzione utilizzando il modello di regressione. Costruire quindi il relativo intervallo di confidenza al livello di confidenza del 95%.

**Esercizio 3**

Su un campione di  $n = 32$  soggetti adulti si sono osservate le seguenti variabili:  $Y$  = qualità della vita (misurata su scala 0-100);  $X_1$  = Anni di istruzione;  $X_2$  = Status occupazionale (0 = Disoccupato; 1 = Occupato). Si supponga di adattare un modello di regressione che pone la qualità della vita,  $Y$ , in funzione delle variabili esplicative  $X_1$  e  $X_2$  e della loro interazione. Si supponga di essere interessati a confrontare tale modello con il modello nullo sapendo la somma dei quadrati dei residui per modello in esame è  $SSE = 430$  e che la varianza della qualità della vita è  $s_Y^2 = 25$

- Completare la seguente tavola di analisi della varianza

Fonte di variabilità	Somma dei Quadrati	Gradi di libertà
Regressione		
Residua		
Totale		

- (b) Scrivere l'ipotesi nulla e l'ipotesi alternativa relativa al confronto tra il modello esteso e il modello nullo. Scrivere quindi la regione critica al livello di significatività  $\alpha = 0.01$
- (c) Calcolare il valore osservato della statistica test e prendere una decisione.

### Domanda 1

Su un campione di soggetti adulti si sono osservate le seguenti variabili:  $Y$  = qualità della vita (misurata su scala 0-100);  $X_1$  = Anni di istruzione  $X_2$  = Status occupazionale (0 = Disoccupato; 1 = Occupato). La tabella che segue riporta le stime dei minimi quadrati dei coefficienti di un modello di regressione che pone la qualità della vita,  $Y$ , in funzione delle variabili esplicative  $X_1$  e  $X_2$  e della loro interazione.

Variabile	Coefficiente	Errore standard
Costante	45.8	10.2
Anni di istruzione	-0.3	0.08
Status occupazionale: Occupato	1.30	0.2
Interazione	2.25	0.5

- (a) Scrivere l'equazione di regressione stimata
- (b) Interpretare i coefficienti del modello
- (c) Calcolare la qualità della vita media per un soggetto occupato con 20 anni di istruzione

### Domanda 2

In uno studio sulla qualità della vita di soggetti in età adulta viene stimato un modello di regressione logistica in cui la qualità della vita,  $Y$  con  $Y = 1$  per soggetti con qualità della vita medio-alta e  $Y = 0$  per soggetti con qualità della vita medio-bassa e viene posto in relazione a  $X$  = numero di anni di istruzione. Di seguito i risultati ottenuti:

```
summary(glm(Y~X, family=binomial(link="logit")))
```

Call:

```
glm(formula = Y ~ X, family = binomial(link = "logit"))
```

Deviance Residuals:

```
Min      1Q  Median      3Q      Max
-1.9916 -1.0676  0.5662  0.9090  1.5392
```

Coefficients:

```
          Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.88957    0.69809  -2.707 0.006794**
X              0.20062    0.05697   3.521 0.000429***
```

---

Signif. codes:

```
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 131.79 on 99 degrees of freedom

Residual deviance: 116.94 on 98 degrees of freedom

AIC: 120.94

Number of Fisher Scoring iterations: 3

- (a) Interpretare il segno del coefficiente relativo agli anni di istruzione
- (b) Interpretare il coefficiente relativo agli anni di istruzione in termini di odds.
- (c) Calcolare la probabilità stimata di avere una qualità della vita medio-alta per un soggetto che ha 15 anni di istruzione